

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-043892  
 (43)Date of publication of application : 18.02.1994

(51)Int.CI. G10L 3/00  
 G10L 3/02

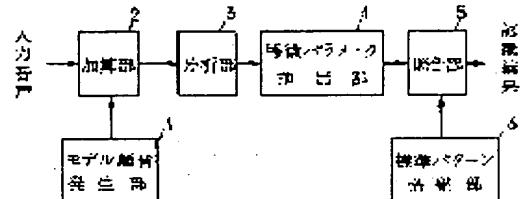
(21)Application number : 04-030395 (71)Applicant : MATSUSHITA ELECTRIC IND CO LTD  
 (22)Date of filing : 18.02.1992 (72)Inventor : FUTAYADA KATSUYUKI  
 MORII TOSHIYUKI

## (54) VOICE RECOGNITION METHOD

### (57)Abstract:

PURPOSE: To provide the voice recognition method which is simple and is superior in a noise resistant characteristic and obtains a stable recognition rate by resolving the problem of a conventional voice recognition device that the recognition rate is reduced by mixture of noise in a use environment or by the noise generated by an input circuit.

CONSTITUTION: A model noise generating part 1 which generates a noise resembling the average features of human voice is provided, and this noise and an input voice are mixed by an adding part 2 to generate the input voice where the noise is mixed, and a feature parameter is obtained by a feature parameter extracting part 4 after analysis of this voice in an analysis part 3. The similarity of the feature parameter to each standard pattern in a standard pattern storage part 6 is obtained by a collating part 5 to output the recognition result.



## LEGAL STATUS

[Date of request for examination] 10.03.1993

[Date of sending the examiner's decision of rejection] 18.07.1995

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-43892

(43)公開日 平成6年(1994)2月18日

(51)Int.Cl.<sup>5</sup>

G 10 L 3/00  
3/02

識別記号 庁内整理番号  
5 2 1 L 7627-5H  
3 0 1 D 7627-5H

F I

技術表示箇所

審査請求 有 請求項の数 4(全 6 頁)

(21)出願番号

特願平4-30395

(22)出願日

平成4年(1992)2月18日

(71)出願人 000005821

松下電器産業株式会社  
大阪府門真市大字門真1006番地

(72)発明者 二矢田 勝行

神奈川県川崎市多摩区東三田3丁目10番1  
号 松下技研株式会社内

(72)発明者 森井 利幸

神奈川県川崎市多摩区東三田3丁目10番1  
号 松下技研株式会社内

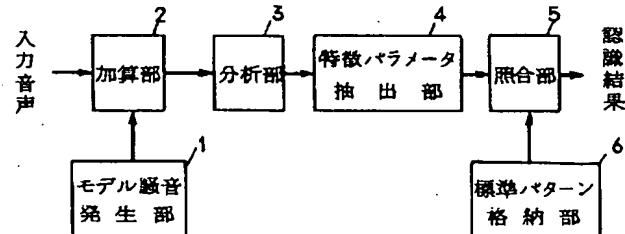
(74)代理人 弁理士 小鍛治 明 (外2名)

(54)【発明の名称】 音声認識方法

(57)【要約】

【目的】 従来の音声認識装置の問題点であった、使用環境での騒音の混入や入力回路で発生するノイズによる認識率の低下といった課題を解決し、簡単な方法によって、耐ノイズ性に優れ、安定した認識率が得られる音声認識方法の提案を目的とする。

【構成】 人の声の平均的な特徴に類似したノイズを発生するモデル騒音発生部1を設け、ここで生成されるノイズを加算部2で入力音声と混合してノイズの混入した入力音声を作成し、これを分析部3で分析した後、特徴パラメータ抽出部4で特徴パラメータを求める。そして、特徴パラメータと標準パターン格納部6の各標準パターンとの類似度を照合部5で求め、認識結果を出力する。



## 【特許請求の範囲】

【請求項1】 音声の平均スペクトルに類似した、性質の明らかなノイズを入力音声に付加し、ノイズを付加した入力音声から得られる特徴パラメータと音声の標準パターンとの照合を行なうことによって音声を認識することを特徴とする音声認識方法。

【請求項2】 音声波形とノイズの波形を混合することによって、ノイズを付加した入力音声を作成することを特徴とする請求項1記載の音声認識方法。

【請求項3】 音声波形を分析して得られるパラメータと、ノイズを分析して得られるパラメータを混合することによって、ノイズを付加した音声入力から得られる特徴パラメータを作成することを特徴とする請求項1記載の音声認識方法。

【請求項4】 入力音声に付加するものと同質のノイズを付加した音声データを用いて音声の標準パターンを作成することを特徴とする請求項1記載の音声認識方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は人間の言葉を機械に理解させるために用いる音声認識方法に関するものである。

## 【0002】

【従来の技術】 音声認識方法の実用化における最大の問題点の1つは耐ノイズ性をいかに向上させるかということである。現在用いられている認識装置や、開発中の認識方法の中には、静かな環境中では十分な認識性能を示すが、騒音が混入すると極端に認識能力が低下するものが少なくない。

【0003】 人間の耳は人の声であるか騒音であるかを正確に聞き分けることができるが、これは物理的な情報のみでなく、言語情報など、より高次の情報をも利用して聴いているためである。物理情報のみを用いる現状の認識装置では、マイクに混入する音声と騒音とを完全に区別して処理するのは不可能であるが、物理情報のみでノイズの影響をできるだけ軽減しようとする試みはいくつかある。

【0004】 騒音には一定レベルの音圧が常に混入する「定常騒音」と、レベルが短時間に大きく変動する「非定常騒音」があるが、ここでは主に前者を問題にする。定常騒音（準定常騒音を含む）の影響を低減する方法として、従来、次の2つがよく用いられている。

【0005】 1つは、長時間平均として求めた騒音スペクトルを、マイクから入力した音声（騒音を含む）から差引く方法である。この方法を用いた第1の従来例として、田部井他「スペクトル・ローカルピークによる単語音声認識」（日本音響学会音声研究会資料 s85-86, Dec. 20, 1985）をあげることができる。この文献では、自動車騒音が混入した音声を11チャンネルの帯域フィルタ

で分析し、各帯域のパワーから騒音スペクトル成分を差引く方法（スペクトルサブトラクション法）によって、騒音の認識に及ぼす影響を軽減している。騒音スペクトルは、認識前に、自動車騒音のみを帯域フィルタで分析して求めておく。

【0006】 音声認識に対する騒音の影響を軽減する第2の方法は、パターンマッチングに用いる標準パターンを、騒音が混入した音声データを用いて作成する方法である。すなわち、入力信号に入る騒音を許容し、そのかわりに、標準パターンに積極的にノイズを混入することによって、スペクトルパターンを適合させようとする方法である。第2の従来例として、発明者らが提案した方法がある（特開昭59-132000号公報「音声の標準パターン作成法」）。これは、一般的な環境騒音を-6dB/カーブのスペクトル傾斜を持つモデル騒音で近似し、これをマイクの入力特性を表現するフィルタに通してノイズデータを作成し、このノイズデータを音声データに対して一定の割合で混入して作成したノイズ入りの音声データを用いて、標準パターンを作成する方法である。

【0007】 この方法は一般的な騒音に対しては有効であり、しかも認識時の処理は全く増えないので、実用的な方法と言える。

## 【0008】

【発明が解決しようとする課題】 入力のスペクトルから騒音のスペクトルを差引く第1の従来例の問題点は、入力信号のレベルが小さいときなどは、差引いた後の値が負になってしまふ危険性があることである。これは、定常騒音と言っても多少の変動があるためである。負になった場合は、差引かないか、あるいは一定値を入れるといった策がとられることがあるが、いずれにしてもスペクトル歪が生じてしまい、認識率に悪い影響がある。また、騒音スペクトルは認識前の短時間で求めるのが普通であるが、この時に、騒音の変動などによって騒音の特徴を正しく求められなかった場合、認識率が大幅に低下してしまう。第1の従来例の方法は、原理的には良さそうな方法であるが、実用として用いる場合にはなかなかうまくいかない。

【0009】 ノイズを混入したデータを用いて標準パターンを作成する第2の従来例の方法では次に示すように、標準パターンに混入したノイズの性質と認識時に混入する騒音の性質が類似している場合は良好な結果になるが、騒音の性質が異なる場合は認識率が低下するという問題がある。（表1）は、日本語の5母音と鼻音のデータを用いて、標準パターン作成用データと評価用データにそれぞれノイズを加えて認識実験を行なった結果である。

## 【0010】

## 【表1】

S N R : 2 0 d B , 単位 %

評価データ＼標準パターン	クリーン	雜踏騒音付加	雨の音付加
クリーン(原音)	<u>9 1. 9 8</u>	9 1. 0 8	8 7. 2 6
雜踏騒音付加	9 0. 1 2	<u>9 1. 1 9</u>	8 6. 4 3
雨の音付加	7 6. 0 6	8 4. 4 1	<u>8 9. 8 5</u>

【0011】ノイズは雜踏騒音（展示会の会場で収録した、人の声を中心とする騒音）と雨の降る音を用い、原音（クリーンデータ）に対して、信号対ノイズ比（S N R）が25dBになるように加えた。（表1）には、標準パターンと評価データに同じノイズを用いた場合、違うノイズを用いた場合、および、原音どうし、原音とノイズ付加の場合の認識率（母音・鼻音の平均認識率）を示している。

【0012】（表1）によると、評価データにノイズが付加している場合は、原音で作成した標準パターンを用いるよりもノイズ付加データで作成した標準パターンを用いる方が結果が良い。そして、標準パターンと評価データのノイズが同じ場合の方が違う場合よりも5%程度認識率が高い（91.19%と86.43%、89.85%と84.41%の比較）。このように第2の従来例は、入力に標準パターンに付加したノイズと同質の騒音が混入した場合には有効な方法であるが、ノイズが異質の場合には良い結果が得られない。

【0013】本発明は従来例における問題点を解決するものである。まず、第1の従来例のような差引く方法を用いないので、騒音の変動に対して影響を受けることはない。そして、第2の従来例に於ける、入力ノイズの性質の違いによる認識率低下という問題点を解決し、優れた音声認識方法を提供するものである。

#### 【0014】

【課題を解決するための手段】本発明は上記課題を解決するために、音声の平均スペクトルに類似した、性質の明らかなノイズを入力音声に付加し、ノイズを付加した入力音声から得られる特徴パラメータと音声の標準パターンとの照合を行なうことによって音声を認識するようにしたものである。

#### 【0015】

【作用】本発明は上記構成により、本来のノイズと後で付加するノイズの両方のノイズ特性が加わった入力音声から得られる特徴パラメータにおける、入力音声に混入する性質が不明の騒音の影響が軽減され、この特徴パラメータと音声の標準パターンとの照合を行なうことで、安定した高い認識率を得ることができる。

#### 【0016】

【実施例】本発明の実施例について図面を用いて詳細に説明する前に、まず、本発明の原理を説明する。前記（表1）において、「雜踏騒音」のスペクトルは多くの人の声の時間平均値とみなして良い。そして「雨の音」のスペクトルは、白色ノイズに近い。（表1）において、原音で作成した標準パターンの項（クリーン）を縦方向に見ると、評価データに雜踏騒音が付加しても認識率の低下はあまり大きくないが（91.98→90.12）、雨の音の騒音が付加すると大きく低下する（91.98→76.06）ことがわかる。すなわち、人の声に類似した騒音の影響は小さいが、人の声と性質が大きく異なる騒音の影響は大きい。

【0017】一般にノイズのスペクトルの傾斜が人の声の平均的なスペクトルの傾きと大きく異なる場合は認識率に対する影響が大きいが、スペクトルの傾きの差異が小さい場合は、認識率にあまり影響を与えない。したがって、人の声に近い性質のノイズならば、音声に重畠しても認識率に与える影響は小さいということである。

【0018】本発明は、このような事実を踏まえ、性質のよく分ったノイズ（人の声に近い性質のノイズ）を入力音声に積極的に加えることによって、マイクから混入する環境騒音や入力系に重畠する電気的ノイズ（白色雑音に近い）など、性質のよく分らないノイズの影響を軽減する方法である。

【0019】図4は、原音データと信号対ノイズ比（S N R）25dB、20dB、15dBの音声データで作成した標準パターンを用いて、原音（クリーン）、S N R 25, 20, 15dBの評価用データ（入力データ）

の認識率を全ての組合わせに対して求めてプロットしたものであり、□印はクリーンデータで作成した標準パターンによる評価を表わし、+印はS N R 25dBのデータで作成した標準パターンによる評価を、△印はS N R 20dBのデータで作成した標準パターンによる評価を、△印はS N R 15dBのデータで作成した標準パターンによる評価をそれぞれ表わしている。付加したノイズは、入力も標準パターンと同じであり、人の声の平均的なスペクトル傾斜に類似した-6dB/オクターブの傾き

を有するモデルノイズである。

【0020】図4によると、標準パターンのS N Rと評価データのS N Rが一致したときに良い結果になっている。そして、両方のS N Rの差が±5 d B程度ならば認識率の低下は小さい。S N R 25 d Bまたは20 d Bの標準パターンを用いれば、クリーンデータからS N R 15 d Bのデータまで、広い範囲の評価データに対して良好な認識率が得られている。従って、入力データのS N Rと標準パターンのS N Rの一一致度は厳密である必要はない。また、図4から、標準パターンは原音で作成しても良いが、それよりも多少のノイズを加えたデータで作成した方が、広い範囲の騒音レベルに対して頑強であることが分る。

【0021】図4の実験は、入力と標準パターンに加えるノイズが全く同じものであった。本発明は、環境ノイズや回路ノイズに対し、性質の分ったノイズを混入する方法があるので、本来のノイズと後で加えるノイズの両

(標準パターン: -6 d B/オクターブノイズ) S N R: 20 d B

道路騒音	工場騒音	学校騒音	-6 d Bノイズ
89.52%	90.57%	90.27%	91.33%

【0024】(表2)のように、入力と標準パターンのノイズが一致している場合(91.33%)に比べ、入力に他の騒音が混入しても認識率の低下はありません。したがって、入力に混入するノイズのスペクトルが少すれても、認識率への影響は小さい。ただし、

(表1)で説明したように、両スペクトルのずれが大きい場合は認識率が大幅に低下する。

【0025】以下、本発明の一実施例について図面を参照して説明する。図1は、前記の手段を用いた本発明の第1の実施例の構成を示すブロック図である。1はノイズ信号を発生するモデル騒音発生部、2は入力信号と前記ノイズ信号を加算する加算部、3はノイズの混入した音声の分析部、4は、分析結果から特徴パラメータを求める特徴パラメータ抽出部、5は特徴パラメータの時系列と標準パターンとのパターンマッチングを行なう照合部、6は音声の標準パターンを格納する標準パターン格納部である。

【0026】上記構成に於てその動作を次に説明する。マイクなどから入力された入力音声(一般に環境騒音や電気ノイズを含む)と、モデル騒音発生部1で発生されたノイズ信号は、加算部2によって加算され、ノイズが混入した入力信号が作成される。モデル騒音は性質がよく分った騒音であり、H o t hスペクトルノイズ、白色雑音を-6 d B/オクターブのフィルタに通して得たノイズ、雜踏騒音の定常部などが用いられる。モデル騒音の加算方法としては、アナログ信号として発生して入力波

方のノイズ特性が入力音声に付加されることになる。したがって、標準パターンに付加するノイズ(性質の分ったノイズ)と異なるノイズが入力に混入することになる。つまり、加えたノイズとは多少異なるノイズが入力に加わる。そこで次に問題になるのは、入力に混入するノイズの性質と標準パターンに混ぜるノイズの性質が、類似してはいるが微妙に異なる場合に、認識率への影響はどの程度かということである。

【0022】これを確かめるために、標準パターンは-10 6 d B/オクターブの特性のモデルノイズを付加して作成し、入力にはこれとは多少スペクトル傾斜が異なるノイズ(道路騒音、工場の騒音、学校の教室で収録した騒音。これらは-6 d B/オクターブに近いスペクトル傾斜を有する)を付加して実験を行なった。結果を(表2)に示す。

【0023】

【表2】

形に直接加える方法、ディジタル信号としてメモリに蓄積しておき、入力音声をAD変換した後で加算する方法、乱数発生器などでランダムノイズを発生して加える方法などがある。入力音声とノイズの割合は、平均的なS N Rが20~25 d B程度にするのがよい。

【0027】分析部3はノイズの混入した音声を分析する部分であり、線形予測分析(L P C分析)を用いているが、その他帯域フィルタ分析、F F T分析などでもよい。特徴パラメータ抽出部4は、分析結果から特徴パラメータを求める部分であり、本実施例ではL P Cケプストラム係数を求めている。分析部3が帯域フィルタの場合には帯域パワーが、F F T分析の場合はF F Tケプストラムや帯域パワーが求められる。照合部5では、入力を分析して得られた特徴パラメータの時系列と標準パターン格納部6に格納されている音声の標準パターンとのパターンマッチングを行ない、類似度が最大になる標準パターンに対応する音声を認識結果として出力する。標準パターンは、認識対象とする各音声に対して、標準パターン作成用データを用いて、あらかじめ作成しておく。また、モデル騒音発生部1で生成されるノイズと同質のものを標準パターン作成用データに一定の割合で加えた後、標準パターンを作成しても良い。パターンマッチングの方法は、入力音声長と標準パターン長を非線形に伸縮して照合する方法や、線形伸縮によって時間長を合せた後照合する方法がある。

【0028】本発明の第2の実施例の構成を示すブロッ

ク図を図2に示す。第1の実施例との違いは、第1の実施例ではモデル騒音発生部1からのノイズ信号を入力音声に加算したが、第2の実施例ではモデル騒音中間パラメータ格納部7を設け、分析部3において、モデル騒音を中間パラメータの形で加えることである。図2において、分析部3でLPC分析を行なうとき、その過程で求められる自己相関関数に対して、モデル騒音中間パラメータ格納部7に格納してあるモデル騒音の自己相関関数を加える方法や、LPC分析の結果として求められる線形予測係数に対して、モデル騒音中間パラメータ格納部7に格納されているモデル騒音の線形予測係数を加工して混合する方法である。特徴パラメータ抽出部4以降の手順は第1の実施例と同様である。

【0029】本発明の第3の実施例の構成を示すブロック図を図3に示す。第3の実施例では、モデル騒音を分析して得たパラメータをモデル騒音パラメータ格納部8に格納しておき、これを特徴パラメータ抽出部4で求めた入力音声の特徴パラメータに対して一定の割合で加える方法である。例えば特徴パラメータとして帯域パワーを用いる場合、入力音声の帯域パワーにノイズの帯域パワーを加えてノイズを混入した帯域パワーを作成し、照合部5へ送る方法が第3の実施例に相当する。

【0030】前記、第1乃至第3の実施例は、いずれも入力音声に対して性質のよく分ったノイズを積極的に混入する方法であり、混入の形態が異なるのみである。第1の実施例が最も精密な方法で効果も大きいが、第2、第3の簡便法でも十分な効果がある。

### 【0031】

【発明の効果】以上述べた本発明を用いることにより、入力音声にマイクから種々の環境騒音が混入した場合や、入力回路の電気的ノイズが重複した場合においても、安定した高い認識率の得られる音声認識装置を作成することが可能となる。本発明を認識装置に組むための処理量の増加や部品点数の増加は極めて少ない。そして、電気的ノイズが多少あってもよいので、安価な入力回路を用いることができるなど、経済的効果も大きい。

10 このように、本発明は音声認識装置の実用化に対しての貢献度が大きく、効果も大きい音声認識方法である。

#### 【図面の簡単な説明】

【図1】本発明の第1の実施例を説明するブロック図

【図2】本発明の第2の実施例を説明するブロック図

【図3】本発明の第3の実施例を説明するブロック図

【図4】評価データの信号対ノイズ比(SNR)と音素認識率の関係を示す概念図

#### 【符号の説明】

- 1 モデル騒音発生部
- 2 加算部
- 3 分析部
- 4 特徴パラメータ抽出部
- 5 照合部
- 6 標準パターン格納部
- 7 モデル騒音中間パラメータ格納部
- 8 モデル騒音パラメータ格納部

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

20

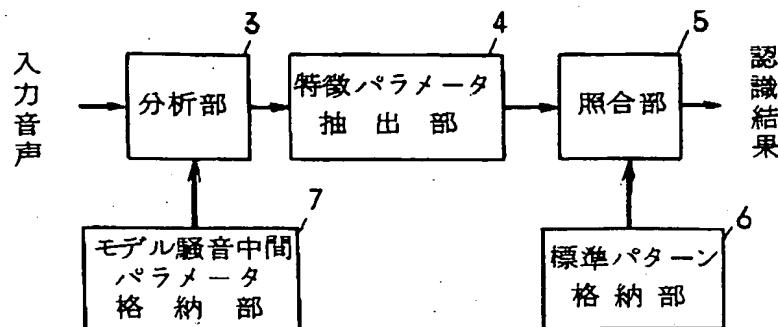
20

20

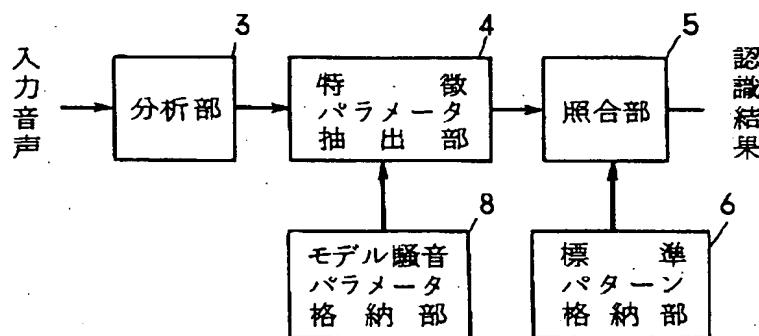
20

20

【図 2】



【図 3】



【図 4】

